



Content Validity and Internal Consistency of a Gordon-Based Household Health Assessment Instrument

Jay Stephen C. Mondong, RN, RM, LPT, MN; Grethel A. Gomez, RM, DM; Honeylyn Villegas, RM; Jofemarie P. Cabal, RM, MPH; Bailey Ray S. Lee, RN, RM, MAN; Keren Happuch O. Noronio, RM

Tangub City Global College

<https://doi.org/10.5281/zenodo.22743414>

ABSTRACT

Household health assessment requires a framework that captures physical, behavioral, psychological, relational, reproductive, coping, and value-based dimensions. Gordon's Functional Health Patterns (GFHP) offers such a structure, but household-level instruments require psychometric evaluation before use. The study aimed to establish the content validity and internal consistency reliability of a GFHP-Based Household Community Needs Assessment Instrument. A methodological, descriptive quantitative design was used. Four experts evaluated content relevance using a four-point scale. Item-Level Content Validity Index (I-CVI), Scale-Level Content Validity Index based on average agreement (S-CVI/Ave), and universal agreement (S-CVI/UA) were computed. The content-validated instrument was pilot-tested with participating household representatives, and Cronbach's alpha was calculated for each GFHP domain and for the complete scale. All evaluated entries received a mean rating of 4.00 and I-CVI = 1.00. S-CVI/Ave and S-CVI/UA were both 1.00. Domain-level alpha coefficients ranged from .72 to .88, indicating acceptable to good internal consistency. Overall alpha was .95 (95% CI [.92, .97]), indicating excellent internal consistency. The instrument demonstrated excellent expert-rated content validity and adequate-to-strong internal consistency in the pilot context. It is appropriate for carefully monitored household assessment while additional dimensionality, stability, criterion-related, and cross-setting evidence is accumulated.

Keywords: *content validity, Cronbach's alpha, Gordon's Functional Health Patterns, household assessment, instrument development, psychometric evaluation*

INTRODUCTION

Health is a multidimensional state encompassing physical, psychological, behavioral, social, and environmental functioning rather than merely the absence of disease (Qadir et al., 2025; Ravaghi et al., 2023). At the household level, these dimensions interact with health-management practices, nutrition, activity, sleep, relationships, reproductive concerns, coping, beliefs, and access to care. Structured household assessment can therefore support community diagnosis, health promotion, referral, and program planning by making these interacting patterns visible.

Gordon's Functional Health Patterns organizes health information into eleven interrelated domains: health perception–health management, nutritional–metabolic, elimination, activity–exercise, sleep–rest, cognitive–perceptual, self-perception–self-concept, role–relationship, sexuality–reproductive, coping–stress tolerance, and value–belief (Gordon, 1994; Hiriscau et al., 2024). GFHP-based tools have supported comprehensive nursing assessment, and prior psychometric studies have shown that structured functional-pattern instruments can produce reliable scores while requiring continued testing across populations and settings (Butcher & Jones, 2021; Butcher et al., 2024; Jones et al., 2012).

Evidence for GFHP application in household and community needs assessment remains less developed than its clinical use. Community assessment instruments frequently emphasize disease, demographic, and environmental indicators, with less systematic coverage of multidimensional functional patterns (Pazzaglia et al., 2023; Ravaghi et al., 2023). A theoretically grounded instrument cannot be considered adequate solely because its content appears appropriate. Expert judgment is needed to establish content relevance and representativeness, followed by empirical reliability analysis to determine whether item responses are sufficiently coherent (Mikkonen et al., 2022; Mokkink et al., 2025).

Research Problem

Household and community assessment requires an instrument that captures multidimensional functioning while producing interpretable and sufficiently consistent scores. Although Gordon's Functional Health Patterns provides a comprehensive conceptual structure, household-level application requires evidence that the instrument content is relevant and representative and that responses within its domains demonstrate adequate internal consistency. The study therefore addressed the need for initial psychometric evidence for a Gordon-based Household Community Needs Assessment Instrument before broader application in community assessment, teaching, research, and program planning.

Assumptions of the Study

The study assumed that the four purposively selected experts possessed sufficient professional competence to judge the relevance and representativeness of the proposed instrument content; that participating household representatives could understand and

respond meaningfully to the content-validated instrument; and that the encoded pilot responses were suitable for estimating internal consistency. Content validity and Cronbach's alpha were treated as initial measurement evidence and not as proof of unidimensionality, temporal stability, criterion validity, or performance across other populations and settings.

Objectives of the Study

The study aimed to develop and psychometrically evaluate a GFHP-Based Household Community Needs Assessment Instrument. It addressed four questions: (1) What was the item-level content validity based on I-CVI? (2) What was the scale-level content validity based on S-CVI/Ave and S-CVI/UA? (3) What was the internal consistency reliability of each GFHP domain? and (4) What was the overall internal consistency reliability of the instrument?

Conceptual or Theoretical Framework

The study was anchored on Gordon's Eleven Functional Health Patterns, which conceptualize health as interrelated patterns of biological, psychological, behavioral, social, relational, and value-based functioning. GFHP provided the substantive structure for instrument development. Content-validity theory and internal-consistency reliability provided the measurement criteria used to evaluate the resulting tool.

Conceptually, the evaluation proceeded through two complementary forms of evidence. Content validity examined whether experts endorsed the relevance and representativeness of the instrument content; internal consistency examined whether responses within each GFHP domain and across the complete scale were coherent. These properties were treated as necessary initial evidence, not as substitutes for construct validity, criterion-related validity, or temporal stability.

METHODOLOGY

Research Design

A methodological research design with a descriptive quantitative approach was used. This design was appropriate because the study developed and evaluated a measurement instrument rather than testing an intervention or causal hypothesis (DeVellis & Thorpe, 2021; Kalkbrenner, 2021). Content validation preceded pilot reliability testing.

Research Locale or Study Setting

The study was conducted in selected communities associated with the extension activities of a Local College in Northern Mindanao, Philippines. The setting corresponded with the instrument's intended household-level application. The pilot was designed for instrument evaluation rather than estimation of community prevalence.

Participants or Data Sources

Two participant groups served distinct psychometric functions. Four experts with relevant competence in midwifery, nursing, public health, community practice, clinical assessment, research, or instrument development were purposively selected for content validation. Accessible adult household representatives from the selected communities were recruited through convenience sampling for pilot reliability testing. This sampling supported preliminary measurement evaluation but not population-level inference.

Research Instruments or Materials

The researcher-developed Household Community Needs Assessment Instrument was organized according to the eleven GFHP domains and designed to capture biological, psychological, social, behavioral, reproductive, coping, and value-related dimensions of household health. A theory-driven development process translated each domain into observable household-level indicators (DeVellis & Thorpe, 2021; Mikkonen et al., 2022). A structured expert form used a four-point relevance scale: 1 = not relevant, 2 = slightly relevant, 3 = relevant, and 4 = highly relevant. Ratings of 3 or 4 were classified as content-valid for CVI calculation.

Validation of Instrument

The four experts independently assessed the relevance and representativeness of the initial content. Ratings were encoded and summarized using I-CVI, S-CVI/Ave, and S-CVI/UA. I-CVI was the proportion of experts assigning a rating of 3 or 4 to an evaluated entry. S-CVI/Ave was the average I-CVI across entries, and S-CVI/UA was the proportion of entries achieving universal agreement. All entries reached the acceptance criterion; no validity-driven deletion was required before pilot testing.

The content-validated instrument was administered to participating household representatives. Responses were encoded by GFHP domain and for the complete instrument. Cronbach's alpha was calculated separately for the eleven domains and for the

overall scale. The 95% confidence interval for the overall alpha was also obtained to represent precision.

Data Gathering Procedure

Data collection occurred in two phases. During Phase 1, the initial instrument and rating form were provided to the experts; completed ratings were checked, encoded, and analyzed. During Phase 2, household participants received information regarding the study and completed the validated instrument after consent. Forms were checked for usable responses, assigned non-identifying codes, encoded by domain, and subjected to reliability analysis.

Data Analysis

I-CVI was calculated as the number of experts assigning a rating of 3 or 4 divided by the total number of experts. S-CVI/Ave was the sum of I-CVIs divided by the number of evaluated entries. S-CVI/UA was the number of entries receiving universal agreement divided by the total number of entries. Scale-level CVI values of .90 or higher were interpreted as excellent. Cronbach's alpha values from .70 to .79 were interpreted as acceptable, .80 to .89 as good, and .90 or higher as excellent. Alpha was treated as evidence of internal consistency, not proof of unidimensionality (Chiavone et al., 2022; Wakelin et al., 2023).

Ethical Considerations

Institutional authorization was secured before data collection. Experts and household participants received information about the study's purpose, procedures, voluntary nature, and right to decline or withdraw. Informed consent was obtained. Non-identifying codes were used, access to data was restricted to the research team, and results were reported in aggregate. No clinical intervention was imposed, and interpretation was limited to the psychometric properties examined.

RESULTS AND DISCUSSION

Results for Objective 1 Item-Level Content Validity

Table No. 1 presents the expert ratings and item-level content validity results.

Entry	Expert 1	Expert 2	Expert 3	Expert 4	Mean	I-CVI	Interpretation
1	4	4	4	4	4.00	1.00	Highly Relevant
2	4	4	4	4	4.00	1.00	Highly Relevant
3	4	4	4	4	4.00	1.00	Highly Relevant
4	4	4	4	4	4.00	1.00	Highly Relevant
5	4	4	4	4	4.00	1.00	Highly Relevant
6	4	4	4	4	4.00	1.00	Highly Relevant
7	4	4	4	4	4.00	1.00	Highly Relevant
8	4	4	4	4	4.00	1.00	Highly Relevant
9	4	4	4	4	4.00	1.00	Highly Relevant
10	4	4	4	4	4.00	1.00	Highly Relevant
11	4	4	4	4	4.00	1.00	Highly Relevant
Total	44	44	44	44	4.00	1.00	—

Table No. 1. Expert Ratings and Item-Level Content Validity

Note. Ratings used a four-point relevance scale. An I-CVI of 1.00 denotes universal agreement among four experts. Expert totals are 44 because each expert assigned 4 to all 11 evaluated entries.

Every evaluated entry received a rating of 4 from all experts. Each entry therefore obtained a mean of 4.00 and I-CVI = 1.00. No expert classified an entry as not relevant, slightly relevant, or merely relevant.

Discussion of Objective 1 Item-Level Content Validity

The I-CVI results demonstrated complete agreement that the represented content was highly relevant to the intended GFHP-based household constructs. The scale-level indices confirmed that this endorsement extended across the full evaluated set rather than selected domains. These findings are consistent with instrument-development guidance emphasizing expert review of relevance, representativeness, comprehensiveness, and comprehensibility before empirical score evaluation (Mikkonen et al., 2022; Mokka et al., 2025).

The findings support the substantive alignment of the instrument with the eleven-pattern framework. However, perfect CVI values should be interpreted specifically as unanimous relevance judgments among the four selected experts. They do not independently establish factor structure, criterion relationships, response-process validity, or performance in other populations.

Results for Objective 2 Scale-Level Content Validity

Table No. 2 presents the scale-level content validity indices.

Measure	Value	Interpretation
S-CVI/Ave	1.00	Excellent
S-CVI/UA	1.00	Excellent

Table No. 2. Scale-Level Content Validity Indices

Note. S-CVI/Ave = average of the I-CVIs; S-CVI/UA = proportion of entries with universal agreement. Values $\geq .90$ indicate excellent content validity.

Both scale-level indices were 1.00. The average item-level agreement reached the maximum possible value, and every evaluated entry achieved universal agreement.

Discussion of Objective 2 Scale-Level Content Validity

The I-CVI results demonstrated complete agreement that the represented content was highly relevant to the intended GFHP-based household constructs. The scale-level indices confirmed that this endorsement extended across the full evaluated set rather than selected domains. These findings are consistent with instrument-development guidance emphasizing expert review of relevance, representativeness, comprehensiveness, and comprehensibility before empirical score evaluation (Mikkonen et al., 2022; Mokkink et al., 2025).

The findings support the substantive alignment of the instrument with the eleven-pattern framework. However, perfect CVI values should be interpreted specifically as unanimous relevance judgments among the four selected experts. They do not independently establish factor structure, criterion relationships, response-process validity, or performance in other populations.

Results for Objective 3 Internal Consistency Across GFHP Domains

Table No. 3 presents internal consistency estimates for the eleven GFHP domains.

GFHP Domain	Cronbach's α	Interpretation
Health Perception–Health Management	.81	Good
Nutritional–Metabolic	.79	Acceptable
Elimination	.79	Acceptable
Activity–Exercise	.83	Good
Sleep–Rest	.79	Acceptable
Cognitive–Perceptual	.88	Good
Self-Perception–Self-Concept	.79	Acceptable
Role–Relationship	.79	Acceptable
Sexuality–Reproductive	.72	Acceptable

GFHP Domain	Cronbach's α	Interpretation
Coping–Stress Tolerance	.81	Good
Value–Belief	.80	Good

Table No. 3. Internal Consistency Across GFHP Domains

Note. Alpha values .70–.79 were acceptable, .80–.89 good, and $\geq .90$ excellent. Value–Belief (.80) follows the stated good-reliability criterion.

Domain coefficients ranged from .72 to .88, and every domain met the minimum .70 criterion. Cognitive–Perceptual had the highest coefficient ($\alpha = .88$). Activity–Exercise obtained $\alpha = .83$; Health Perception–Health Management and Coping–Stress Tolerance each obtained $\alpha = .81$; and Value–Belief obtained $\alpha = .80$. Sexuality–Reproductive had the lowest coefficient ($\alpha = .72$) but remained acceptable.

Discussion of Objective 3 Internal Consistency Across GFHP Domains

Acceptable-to-good alpha coefficients across all domains indicate that internal coherence was not confined to the complete scale. This is important for a multidomain instrument: users may need to interpret household needs by functional pattern, and adequate within-domain reliability is therefore necessary. The observed pattern is compatible with prior GFHP instrument research showing that structured functional-pattern content can produce satisfactory reliability while requiring continued testing and cultural refinement (Butcher et al., 2024; Jones et al., 2012).

The Sexuality–Reproductive Pattern obtained the lowest coefficient. Its $\alpha = .72$ remained acceptable but suggests comparatively greater response variation. Possible contributors include household composition, item applicability, cultural sensitivity, privacy, and reluctance to disclose reproductive or sexual-health information. These explanations were not directly tested and should guide cognitive interviewing and item-level analysis rather than be treated as demonstrated causes.

Results for Objective 4 Overall Internal Consistency

Table No. 4 presents the internal consistency estimate for the complete instrument.

Scale	Cronbach's α	95% CI	Interpretation
Overall instrument	.95	[.92, .97]	Excellent

Table No. 4. Overall Internal Consistency

Note. CI = confidence interval. Alpha $\geq .90$ indicates excellent internal consistency.

The overall alpha was .95, with a 95% confidence interval from .92 to .97. Both interval limits remained within the excellent range.

Discussion of Objective 4 Overall Internal Consistency

The overall $\alpha = .95$ and narrow confidence interval indicate excellent internal consistency for the complete instrument in the pilot data. The estimate complements, rather than replaces,

the domain coefficients: the overall value reflects coherence across the instrument, whereas domain values reflect consistency within theoretically distinct patterns.

A coefficient of .95 also warrants examination for item overlap. Because GFHP domains are conceptually interrelated, shared variance across items is expected; nevertheless, corrected item–total correlations, inter-item correlations, alpha-if-item-deleted statistics, and factor-analytic evidence are needed to determine whether any content is redundant. Alpha alone cannot demonstrate unidimensionality or that all items are indispensable.

General Discussion

Taken together, the findings provide strong initial evidence for the instrument within the pilot context. Complete expert agreement produced I-CVI, S-CVI/Ave, and S-CVI/UA values of 1.00, while all eleven GFHP domains met the minimum internal-consistency criterion and the complete scale obtained excellent reliability. The results support substantive alignment with Gordon’s framework and coherent scoring, while also defining the boundaries of the evidence: content validity and internal consistency do not independently establish dimensionality, criterion relationships, response-process validity, temporal stability, or generalizability across settings.

The instrument can support supervised household assessment by midwives, community health practitioners, students, and faculty in settings comparable to the pilot context. It offers a structured way to complement disease, demographic, environmental, and service-utilization data with functional-pattern information. Aggregated results may inform community diagnosis, health teaching, referral priorities, extension activities, and program planning. Use should remain within the administrator’s scope of practice and should complement established clinical-assessment and emergency-referral pathways.

Institutional users may adopt the instrument for supervised community exposure and research after standardized orientation on administration, confidentiality, scoring, referral, and handling of sensitive responses. A user manual should define respondent eligibility, non-applicable items, missing data, domain scoring, interpretation, and urgent-referral procedures.

Future studies may use cognitive interviews and item-level diagnostics, particularly for the Sexuality–Reproductive domain and possible overlap suggested by the overall alpha. Larger and more diverse samples can extend the evidence through factor analysis, test–retest reliability, inter-rater reliability where applicable, convergent or criterion-related validity, measurement invariance, and cross-setting performance. Substantive revisions should be followed by renewed content validation and reliability testing.

Limitations

The evidence represents an initial, context-specific validation phase. Purposive expert selection and convenience recruitment of household representatives were suited to the study’s preliminary measurement purpose, although findings should be applied primarily in

settings comparable to the pilot context. The study examined content validity and internal consistency; subsequent research can extend the evidence to temporal stability, inter-rater agreement, criterion relationships, structural validity, measurement invariance, and responsiveness to change. These boundaries define the present scope while retaining the strong content-validity and reliability results obtained.

The study focused on development and initial psychometric evaluation. Four purposively selected experts provided content-validity ratings, and conveniently recruited household representatives from communities associated with the college's extension activities supplied pilot reliability data. The evidence pertains to the eleven GFHP domains and to the participating pilot context.

CONCLUSION

The GFHP-Based Household Community Needs Assessment Instrument demonstrated excellent expert-rated content validity, acceptable-to-good internal consistency across all eleven GFHP domains, and excellent overall internal consistency. These findings support its initial use for carefully monitored household assessment in contexts comparable to the pilot setting. Continued evaluation across larger and more diverse samples can progressively extend the evidence for its broader application.

DECLARATIONS

Ethics approval and consent to participate. Institutional authorization was secured before data collection. Participation was voluntary, informed consent was obtained, confidentiality was maintained through non-identifying codes and restricted data access, and findings were reported in aggregate.

Consent for publication. Not applicable because no individually identifiable information is presented.

Data availability. De-identified data may be made available by the corresponding author upon reasonable request, subject to institutional authorization and applicable privacy requirements.

Competing interests. The authors are affiliated with the institution involved in the instrument's development and pilot use. This relationship is disclosed as a potential institutional interest; affiliations are omitted from the anonymized manuscript.

Funding. No external funding was reported for the study.

REFERENCES

- Banister, G., Carroll, D. L., Dickins, K., Flanagan, J., Jones, D., Looby, S. E., & Cahill, J. E. (2022). Nurse-sensitive indicators during COVID-19. *International Journal of Nursing Knowledge*, 33(4), 234–244. <https://doi.org/10.1111/2047-3095.12372>
- Bitencourt, G. R., Souza, P. A., Ferreira, A. F. M., Fernandes, L. L. R. A., Silva, C. S., Souza, P. S., & Corrêa, D. B. S. F. (2023). Functional patterns of health nursing theory in the hospital context: Evaluation according to Meleis. *Global Academic Nursing Journal*, 4(1), e336. <https://doi.org/10.5935/2675-5602.20200336>
- Butcher, R. C. G. S., Guandalini, L. S., Barros, A. L. B. L., Damiani, B. B., & Jones, D. A. (2024). Psychometric evaluation of the Functional Health Pattern Assessment Screening Tool – Modified Brazilian version. *Revista Latino-Americana de Enfermagem*, 32, e4119. <https://doi.org/10.1590/1518-8345.6755.4119>
- Butcher, R. C., & Jones, D. A. (2021). An integrative review of comprehensive nursing assessment tools developed based on Gordon's eleven functional health patterns. *International Journal of Nursing Knowledge*, 32(4), 294–307. <https://doi.org/10.1111/2047-3095.12321>
- Chiavone, F. B. T., Andrade, F. B., Silva, A. F. M., Azevedo, I. C., Martins, Q. C. S., & Santos, V. E. P. (2022). Types of analysis of validation studies in nursing: A scoping review. *Investigación y Educación en Enfermería*, 40(3), e09. <https://doi.org/10.17533/udea.iee.v40n3e09>
- DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). SAGE Publications.
- Gordon, M. (1994). *Nursing diagnosis: Process and application* (3rd ed.). Mosby.
- Govindasamy, P., Cumming, T. M., & Abdullah, N. (2024). Validity and reliability of a needs analysis questionnaire for the development of a creativity module. *Journal of Research in Special Educational Needs*, 24, 637–652. <https://doi.org/10.1111/1471-3802.12659>
- Hiriscu, E. I., Cauli, O., Donca, V., Marinescu, L.-A., Macarie, A.-E., Avram, L., Cancel, O.-G., Donca, S., Buzdugan, E.-C., Crisan, D.-A., & Bodolea, C. (2024). The association between functional health patterns and frailty in hospitalized geriatric patients. *Geriatrics*, 9(2), 41. <https://doi.org/10.3390/geriatrics9020041>
- Jones, D., Duffy, M. E., Flanagan, J., & Foster, F. (2012). Psychometric evaluation of the Functional Health Pattern Assessment Screening Tool (FHPAST). *International Journal of Nursing Knowledge*, 23(3), 140–145. <https://doi.org/10.1111/j.2047-3095.2012.01224.x>
- Kalkbrenner, M. T. (2021). A practical guide to instrument development and score validation in the social sciences: The MEASURE approach. *Practical Assessment, Research, and Evaluation*, 26(1), 1–17. <https://doi.org/10.7275/svg4-e671>
- Kim, T. Y., Baldrias, L., Papageorgiou, S., et al. (2022). A community-based survey to assess risk for one health challenges in rural Philippines. *One Health Outlook*, 4, 7. <https://doi.org/10.1186/s42522-022-00063-0>
- Kobashi, Y. (2025). Health needs assessment tool for identifying health issues among community residents with unmet needs. *Tropical Medicine and Health*, 53, 29. <https://doi.org/10.1186/s41182-025-00713-9>
- Mikkonen, K., Tomietto, M., & Watson, R. (2022). Instrument development and psychometric testing in nursing education research. *Nurse Education Today*, 119, 105603. <https://doi.org/10.1016/j.nedt.2022.105603>

- Mokkink, L. B., Herbelet, S., Tuinman, P. R., & Terwee, C. B. (2025). Content validity: Judging the relevance, comprehensiveness, and comprehensibility of an outcome measurement instrument—A COSMIN perspective. *Journal of Clinical Epidemiology*, 185, 111879. <https://doi.org/10.1016/j.jclinepi.2025.111879>
- Pazzaglia, C., Camedda, C., Ugenti, N. V., et al. (2023). Community health assessment tools adoptable in nursing practice: A scoping review. *International Journal of Environmental Research and Public Health*, 20, 1667. <https://doi.org/10.3390/ijerph20031667>
- Qadir, J., Maddah, D., Qoronfleh, M. W., & Senturk, R. (2025). Toward multiplex health: Integrating complexity, normativity, and open science. *Frontiers in Psychology*, 16, 1603474. <https://doi.org/10.3389/fpsyg.2025.1603474>
- Rahmat, R. A., Saidi, S. B., & Nasir, N. S. M. (2024). Content validity of digital knowledge using CVI method. *Environment-Behaviour Proceedings Journal*, 9(SI20), 21–28. <https://doi.org/10.21834/e-bpj.v9iSI20.6092>
- Ravaghi, H., Guisset, A.-L., Elfeky, S., et al. (2023). A scoping review of community health needs and assets assessment. *BMC Health Services Research*, 23, 44. <https://doi.org/10.1186/s12913-022-08983-3>
- Sabo, A., Kuan, G., Abdullah, S., Kuay, H. S., Goni, M. D., & Kueh, Y. C. (2024). Psychometric properties of the social determinants of health questionnaire (SDH-Q): Development and validation. *BMC Public Health*, 24, 2507. <https://doi.org/10.1186/s12889-024-19990-w>
- Sayili, U., Sak, K., Aydin, S. N., et al. (2024). Development, validity and reliability of the healthy lifestyle behavior scale. *Discover Public Health*, 21, 62. <https://doi.org/10.1186/s12982-024-00186-x>
- Wakelin, K. J., McAra-Couper, J., Fleming, T., & Erlam, G. D. (2023). Assessing reliability and validity of online survey questions. *Methodological Innovations*, 16(1), 91–101. <https://doi.org/10.1177/20597991221148401>